PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
G06F 13/362, 13/40

A1

(11) International Publication Number: WO 96/38792

(43) International Publication Date: 5 December 1996 (05.12.96)

(21) International Application Number: PCT/EP96/02338

(22) International Filing Date: 30 May 1996 (30.05.96)

(30) Priority Data:
9510935.1 31 May 1995 (31.05.95) GB

(71) Applicant *(for all designated States except US)*: 3COM IRELAND [-/-]; Upland House, P.O. Box 309, Georgetown, Grand Cayman (KY).

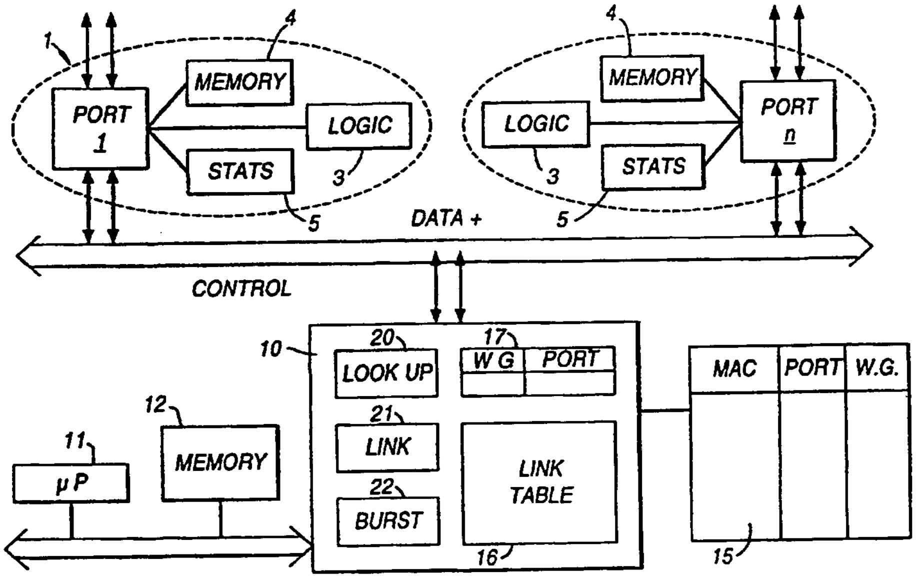
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: CREEDON, Tadhg [IE/IE]; Coismeagmore, Furbo, County Galway (IE). O'CONNELL, Anne [IE/IE]; 3 Woodberry, Carpenterstown Road, Castleknock, Dublin 15 (IE). O'NEILL, Eugene [IE/IE]; 18 Stillgorn Heath, Upper Kilmacud Road, County Dublin (IE). GAVIN, Vincent [IE/IE]; 18 The Old Rectory, Chapel Hill, Lucan, County Dublin (IE). HICKEY, John [IE/IE]; Rathmoley, Killenaule, County Tipperary (IE). GAHAN, Richard [IE/IE]; Kilcoulshea, Ferns, County Wexford (IE). SHERER, William, Paul [IE/US]; 1054 Gardenia Way, Sunnyvale, CA 94086 (US).

(74) Agent: CRAWFORD, Andrew, Birkby; A.A. Thornton & Co., Northumberland House, 303-306 High Holborn, London WC1V 7LE (GB).

(81) Designated States: AU, CA, GB, JP, KR, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: MONITORING AND CONTROL OF DATA FLOW IN A COMPUTER NETWORK DEVICE



(57) Abstract

Apparatus for monitoring and controlling data flow in a computer network device having a plurality of parts comprises control means for directly linking ports together on the basis of additional information stored in the control means whereby incoming packets are linked directly to an output port to achieve high performance. The additional information is stored in one or more look-up tables additional to the normal CAM with the or each table addressed by separate processing. This allows the implementation to be in hardware rather than in software.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia
AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgystan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

MONITORING AND CONTROL OF DATA FLOW IN A COMPUTER NETWORK DEVICE

The present invention relates to apparatus for monitoring and controlling data flow in a computer network device.

Computer networks are well known and basically there are two popular types i.e. a token ring network and an ethernet network. Such networks are now well defined. In other words, in order to have compatible equipment and software certain features have to be present in order to comply with the standard.

It is known to divide up a large network using devices called bridges and in ethernet technology a bridge is a defined device having defined characteristics. However, we have in the past modified such devices as that they retain the defined operations of a bridge but also, within the device, handle data in a different manner so as to economize on memory. In our terminology a modified bridge is termed a switch.

In modern networks, more and more control of data flowing in the network is required in order to avoid bottle-necks which cause delays. There is thus a need for a high performance, low cost switch or bridge.

The present invention proposes apparatus for monitoring and controlling data flow in a computer network device having a plurality of ports, the apparatus including control means for directly linking ports together on the basis of additional information stored in the control means whereby incoming packets are linked directly to an output port to achieve high performance.

Preferably, the additional information is stored in look-up table means additional to the normal CAM or equivalent mechanism. The look-up table means could be in the form of one large table or a plurality of smaller tables. Each table is addressed using separate processing.

This arrangement with its system of tables is ideal for implementation in hardware (e.g. in silicon), rather than in software, thus allowing for low cost implementation.

In order that the present invention be more readily understood an embodiment thereof will now be described by way of example with reference to accompanying drawings, in which:-

Fig. 1 shows a diagrammatic lay out of the structure of a device according to the present invention;

Fig. 2 shows a representation of one part of a device as shown in Fig. 1; and

Fig. 3 shows diagrammatically a typical burst of data for transmission between ports in the device shown in Fig. 1.

The preferred embodiment of the present invention is shown in Fig. 1 and will be seen to comprise a multi-port switch having ports 0 to n to which devices such as computer terminals, servers, printers and modems may be attached.

Within the switch there is a data bus and a control bus, although for convenience, only a single bus is indicated in the drawing. Each port is connected to the data and control buses. It is common to select one particular port to connect to rest of network. This port is known as the downlink port.

As will be seen from Fig.1, each port is provided with transmit and receive buffers 4 in the form of memory devices.

Traffic through the switch is controlled by a control device generally indicated by the reference numeral 10 which will be described in detail later. The control device is connected to the data and control buses and also communicates with a switch management entity in the form of a processor 11 and memory 12.

The control device 10 will now be described in more detail. It consists of look-up tables which are written into and read from under the control of three separate processing elements. One of the look-up tables is basically a modified content addressable memory (CAM) 15 or equivalent mechanism for storing MAC addresses and associated port numbers in a conventional manner. The CAM is used to store information associated with each MAC address, such as port number, age, type etc. Operation of the control device 10 is based upon a link table 16 which contains information relating to each of the ports of the switch and this table 16 is shown in more detail in Fig. 2. There is also a small down link table 17 and the inter-relationship between the three tables 15, 16 and 17 will also be described.

Let us assume there are 29 ports in the switch and that the device attached to port 0 wishes to communicate with another device in the network.

As soon as the packet stated being received by port 0, a lookup request flag is written into the memory section of port 0.

The lookup machine 20 in the control device 10 scans each receive port, whose port enable bit in the link table is set, for lookup request flags. On finding a lookup request, the lookup process is carried out by the lookup engine. This causes the source and destination addresses to be read from port 0 and stored by the control device. The lookup engine will determine the destination port for the packet, and write the destination port, a multicast/broadcast indicator and a link request flag into the link table. It also causes the lookup request flag to be cleared in port 0.

The link engine, on finding the link request flag set in the link table, will decide if the destination port is busy or not. If not, it sets a burst request flag in the link table. If it is busy, it does not set the bit in the link table, but sets a flag in the store Rx column and causes the packet to be stored at port 0 until the destination port becomes free, whereupon the link engine will then set the burst request flag.

The burst engine, on seeing the burst request flag bit set, then permits port 0 to transmit to the destination port via the switch data bus the packet stored in its transmit memory. The preferred form of the burst is shown in Figure 3.

Once the transmission is concluded, which is determined by the burst engine, the link engine clears the task table of the entries relating to the source port in question. After each engine is finished its link it moves on to the next port. It is thus apparent that all processing engines are independent but act sequentially by passing a request on to the next engine by placing flags at the appropriate location in the link table. Also, the packet is not sent to a central store. Rather, it remains at the source port and, under the control of the burst engine, passes along the switch bus once to the destination port.

While the basic operation described above relates to the situation where both the source MAC address and destination MAC address are known to the control unit 10, it also applies to situations where the destination address is not known. This is because all unknown destination addresses are sent to the downlink port. The control unit keeps track of device MAC addresses and corresponding ports using a CAM in a conventional manner although here it is proposed that all destination addresses not on a port of the switch will be designated as having the downlink port as their associated port.

The above description applies to unicast traffic. In the case of broadcast or multicast traffic the operation is slightly different. For a port wishing to send a broadcast or multicast packet, this request is noted by the look-up engine 20 which sets not only a link request flag but also a multicast/broadcast request flag. Once all destination ports are ready to receive the packet, the link engine sets the burst flag and the burst engine then causes the packet to be transmitted once on to the switch data bus from where it is received simultaneously by all the destination ports.

Once the broadcast has taken place, the burst request having been removed, the link engine clears the link table for that entry.

The link table is also provided with indications for "cut through" operation. The link engine 21 can determine whether cut-through operation is appropriate having regard to the source and destination ports and if it is appropriate it places a flag in the cut-through column.

As mentioned above, the link table has a column indicating whether or not the requested transmission is of a broadcast or multicast packet. To facilitate handling of such requests, we propose that the ports of the switch be allocated an additional address indicating that devices attached thereto should be grouped together for operational purposes. These groups of device we will call work groups (WG).

Each port is given a work group number and as the device attached to a port wishes to communicate with another device, the packet has associated with it the work group number allocated to the port to which the source device is connected. When the packet is transmitted across the switch bus, the work group of the source port is also transmitted as shown in Figure 3. For multicast/unicast packets, only those destination ports whose work group matches will accept the packet. For unicast packets, only the destination port which matches the destination port and work group on the switch bus will accept the packet.

The use of work groups enables virtual networks to be set up within one hardware network. This avoids devices not in the work groups having to deal with broadcast messages not of interest to them.

CLAIMS:

1. A computer network device comprising

a plurality of ports each arranged to be connected to external apparatus, and

a bus coupling the ports together in order to provide data flow between selected ports wherein

each port is provided with storage means for storing data to be transmitted to another port via the data bus and for indicating the desire for such transmision and wherein

a control device is provided which comprises means for recognising the desire of a port to transmit data, means for determining the destination port of the transmitted data and means for controlling access of the port to the bus for transmission of the data until such time as the destination port is available to receive such data.

2. A computer network device of claim 1, wherein the control device includes look-up means for monitoring the ports for the indication of the desire to transmit data and generating a transmit request flag, link means responsive to the presence of a transmit request flag for storing the destination port or ports to which data should be transmitted, and further means for monitoring the status of destination ports and for permitting transmission of the data on the bus to the or each destination port.

3. A computer network device of claim 2, wherein the further means is arranged to permit transmission of data for a predetermined period of time.

4. A computer network device of claim 3, wherein the further means is arranged to permit a plurality of

transmissions of data, each for said predetermined period of time, until all data stored at the transmitting port has been transitted.

5. A computer network device of claims 2, 3 or 4, wherein the look-up means, link means and further means are arranged to operate sequentially, and once having completed a task, return to an initial state.

6. A computer network device of any of claims any 2 to 5 and comprising a look-up table for storing the results of the operation of the look-up means, the link means and the further means.

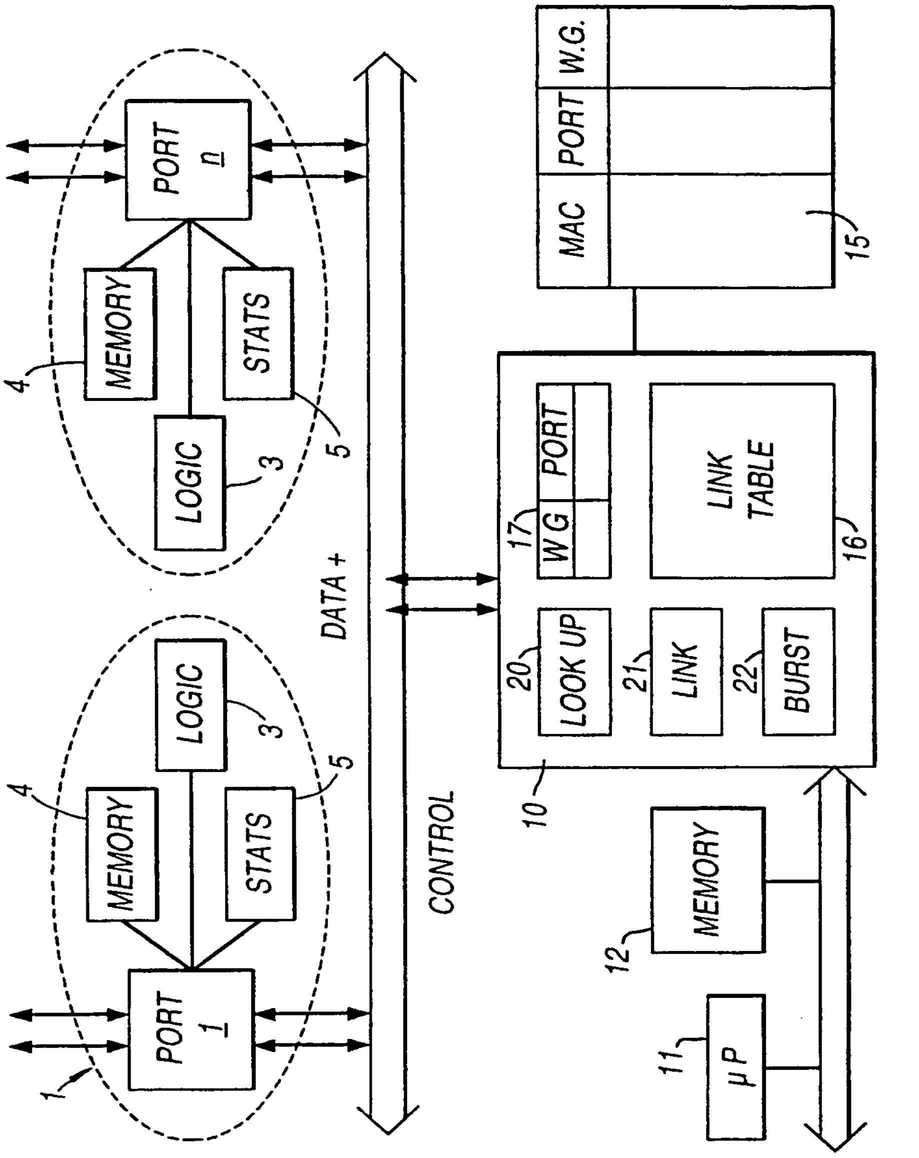

PORT
n
MEMORY
STATS
LOGIC
4
3
5
PORT
1
MEMORY
STATS
LOGIC
1
4
3
5
DATA +
CONTROL
MAC
PORT
W.G.
15
W G
PORT
17
LINK
TABLE
16
LOOK UP
20
LINK
21
BURST
22
10
MEMORY
12
μP
11

Fig. 1

	1 bit	1 bit	6 bit	1 bit	1 bit	1 bit	1 bit	3 bit
	LOOK UP REQ	LINK ENABLE	DEST PORT	MULTI CAST/ BROADCAST PKT	CUT THRU	PORT ENABLE	STORE R X	OTHER CONT'L
Ø	1	0	12	0	0	1	1	....
1								
....								
12								

Fig.2

DEST. PORT # W.G. CONTROL	DATA	DATA	DATA	DATA

Fig.3

INTERNATIONAL SEARCH REPORT

Internationa plication No
PCT/EP 96/02338

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G06F13/362 G06F13/40

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G06F H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO,A,93 03439 (TANDEM COMPUTERS INCORPORATED) 18 February 1993 see abstract see page 7, line 34 - page 10, line 2 see claims 1,11-13; figures 1-3 ---	1
A	US,A,5 140 585 (TOMIKAWA) 18 August 1992 see abstract see column 1, line 64 - column 2, line 33 see column 3, line 27 - column 5, line 8 see figures 1,3 ---	1-6
A	EP,A,0 556 148 (DIGITAL EQUIPMENT CORPORATION) 18 August 1993 see column 4, line 15 - column 7, line 36 see claims 1,2; figure 3 -----	1-6

□ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search 11 October 1996	Date of mailing of the international search report 07. 11. 96
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+ 31-70) 340-3016	Authorized officer McDonagh, F

Information on patent family members

International Application No
PCT/EP 96/02338

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9303439	18-02-93	EP-A-	0597013	18-05-94
		US-A-	5455917	03-10-95
US-A-5140585	18-08-92	JP-A-	4079442	12-03-92
		JP-A-	4079443	12-03-92
EP-A-556148	18-08-93	JP-A-	6085819	25-03-94
		US-A-	5524254	04-06-96